Effects of discourse factors on prosodic rhythm:

Evidence from inter- and intra-speaker variation in Hispanic English

Charlotte Vaughn
Northwestern University
crvaughn@northwestern.edu

N WAV37 | Rice University | Houston, TX | November 8, 2008
Rhythm

• Languages are often classified as more **stress-timed**, e.g. English, Dutch, Arabic; or more **syllable-timed**, e.g. Spanish, Italian, Mandarin (Abercrombie 1967, Pike 1945)
  • Or mora-timed, e.g. Japanese

• Many attempts to find phonetic correlates of rhythm and a quantification metric (e.g. Dauer 1983, Ramus et al. 1999; Low et al. 1995)
Pairwise Variability Index (PVI)

- Low and Grabe (1995)
- Low, Grabe and Nolan (2000)
- Thomas and Carter (2006)
  (PVI as used here ≈ nPVI or nPVI-V)

\[ PVI = \frac{\text{absolute value of } V_1 - V_2}{\text{mean of } V_1 \text{ and } V_2} \]

Difference between lengths of two adjacent vowels

Normalizes for changes in speech rate over time
Previous PVI studies

• Used aggregate PVI scores to represent whole languages or varieties
  – Low et al. 2000 – Singapore English/British English
  – Spencelayh 2001 – communities of British English speakers
  – Grabe and Low 2002 – typology of world languages
  – Fought and Fought 2002; Carter 2004; Carter 2005 – Chicano/Hispanic English
  – Thomas and Carter 2006 – African American English
  – Szakay 2006 – New Zealand English
  – Coggshall 2007 - Lumbee English and Cherokee
  – Ghazali, Hamdi and Barkat 2002 – dialects of Arabic

• Often look at substrate effects
“The figure shows that the differences between individual Spanish speakers are at least as great as the differences between some languages.”

(Grabe 2002)
Intra-speaker variation

“... is rhythm a variable phenomenon like so many of the features found in interlanguage or second language varieties? In other words, might a speaker sound very syllable-timed in some contexts, and very stress-timed in others, or is rhythm more of an immutable trait that is more impervious to variation?” (Carter 2005)

• Intra-speaker variation ≈ stylistic variation (?)
Predictions
(assuming variability occurs)

• Speakers’ PVI will vary significantly depending on their language variety

• A speaker’s PVI will not vary significantly depending on discursive factors
Speech sample

- Interview conducted in 2006 at Chewning Middle School in Durham, North Carolina
- Part of a larger project by the North Carolina Language and Life Project (NCLLP) studying Hispanic English

Two interviewees
- Paco (PM), age 15, male, Hispanic, L2 English, LOR ~6 years
- Tobi (TG), age 15, male, Hispanic, L2 English, LOR ~6 years

Two interviewers
- Danica (DC), age 28, female, white, L1 English
- Erin (EC), age 29, female, white, L1 English

- ~ 30 minute interview, all vowel durations measured for all speakers for last 15 minutes of interview
Descriptive statistics by speaker

<table>
<thead>
<tr>
<th>Speaker</th>
<th>n Vowel Durations</th>
<th>Median PVI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Including Spanish</td>
<td>1104</td>
<td><strong>0.375</strong></td>
</tr>
<tr>
<td>English only</td>
<td>1060</td>
<td><strong>0.380</strong></td>
</tr>
<tr>
<td><strong>TG</strong></td>
<td>538</td>
<td><strong>0.296</strong></td>
</tr>
<tr>
<td><strong>DC</strong></td>
<td>925</td>
<td><strong>0.373</strong></td>
</tr>
<tr>
<td><strong>EC</strong></td>
<td>Including Spanish</td>
<td>556</td>
</tr>
<tr>
<td>English only</td>
<td>468</td>
<td><strong>0.319</strong></td>
</tr>
</tbody>
</table>

Interviewees: **Paco (PM)**, **Tobi (TG)**
Interviewers: **Danica (DC)**, **Erin (EC)**

Speakers’ median PVI ranking, highest to lowest:

- PM
- DC
- EC
- TG
Comparison with other results from Thomas and Carter 2006

Interviewees:
- Paco (PM)
- Tobi (TG)

Interviewers:
- Danica (DC)
- Erin (EC)
Comparison with other results
from Thomas and Carter 2006

Interviewees:
- Paco (PM)
- Tobi (TG)

Interviewers:
- Danica (DC)
- Erin (EC)
Analysis of *inter*-speaker variation

- Why might the interviewers (*DC* and *EC*) have low PVI medians relative to their ethnicity?
  - Age
  - “Good interviewers” accommodating to perceived norm
  - Tendency to speak more “carefully” in interviews
    - Careful enunciation keeps unstressed syllable longer
  - High familiarity with and fluency in Romance language(s)
Analysis of *inter*-speaker variation

- Why might interviewees (PM and TG) have PVI medians relatively spread out from each other? Why might TG’s be lower than PM’s?
  - Different acquisition targets?
  - Carter 2004 discusses possibility of regional and even community-based targets (perhaps simultaneously)
    - PM is in a gang, TG is not; although it is a Hispanic gang, might PM’s identity model be closer to American-ness and thus European American or African American speech?
Results
(after observing variability)

• Speakers’ PVI **will** vary significantly depending on their *language variety*
  - Not in this data set

• A speaker’s PVI **will not** vary significantly depending on *discursive factors*
Examining *intra*-speaker variation

Ways to look at changes during discourse:
- Divide data into *segments* of equal length
- Divide data into *topics* of discussion
- Other variables like speech rate, pause length, *utterance length*
## Distribution of topics over segments

<table>
<thead>
<tr>
<th>Seconds</th>
<th>Topic</th>
<th>PM</th>
<th>TG</th>
<th>DC</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>900</td>
<td>Gangs1</td>
<td>Segment 1</td>
<td></td>
<td>Segment 1</td>
<td>Segment 1</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>Music</td>
<td>Segment 2</td>
<td></td>
<td>Segment 2</td>
<td>Segment 1</td>
</tr>
<tr>
<td>1200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1300</td>
<td>Spanish</td>
<td>Segment 3</td>
<td></td>
<td>Segment 3</td>
<td>Segment 3</td>
</tr>
<tr>
<td>1400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500</td>
<td>Movies</td>
<td></td>
<td></td>
<td></td>
<td>Segment 2</td>
</tr>
<tr>
<td>1600</td>
<td>Break</td>
<td>Segment 4</td>
<td></td>
<td>Segment 2</td>
<td>Segment 4</td>
</tr>
<tr>
<td>1700</td>
<td>Gangs2</td>
<td></td>
<td></td>
<td></td>
<td>Segment 2</td>
</tr>
</tbody>
</table>

**Interviewees:**
- **Paco (PM)**
- **Tobi (TG)**

**Interviewers:**
- **Danica (DC)**
- **Erin (EC)**
### Median PVI per segment

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Segment</th>
<th>n PVI</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PM</strong></td>
<td>1</td>
<td>200</td>
<td>0.394</td>
</tr>
<tr>
<td>Overall Med = 0.380</td>
<td>2</td>
<td>200</td>
<td>0.413</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>200</td>
<td>0.373</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>85</td>
<td><strong>0.286</strong></td>
</tr>
<tr>
<td><strong>TG</strong></td>
<td>1</td>
<td>200</td>
<td>0.339</td>
</tr>
<tr>
<td>Overall Med = 0.296</td>
<td>2</td>
<td>135</td>
<td>0.270</td>
</tr>
<tr>
<td><strong>DC</strong></td>
<td>1</td>
<td>200</td>
<td><strong>0.487</strong></td>
</tr>
<tr>
<td>Overall Med = 0.373</td>
<td>2</td>
<td>200</td>
<td>0.354</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>200</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>84</td>
<td>0.309</td>
</tr>
<tr>
<td><strong>EC</strong></td>
<td>1</td>
<td>200</td>
<td>0.336</td>
</tr>
<tr>
<td>Overall Med = 0.319</td>
<td>2</td>
<td>83</td>
<td>0.308</td>
</tr>
</tbody>
</table>

*Interviewees: Paco (PM), Tobi (TG)*
*Interviewers: Danica (DC), Erin (EC)*

$p = .010$ (comp. to Seg. 1)
$p = .006$ (comp. to Seg. 2)
$p = .012$ (comp. to Seg. 3)
Topic

- Already, we see variation across some of the data during the course of the interview.
- What is causing the first part of DC’s data and the last part of PM’s data to behave differently than the rest?
  - Is it where they fall in the interview?
  - Is it what is being discussed during that segment?
Interview grouped into topics

- **Gangs1**: primarily PM discussing his gang involvement
- **Music**: primarily PM discussing his favorite music
- **Spanish**: EC and PM discuss gang’s Spanish terms
- **Movies**: PM and TG talk about upcoming movies
- **Break**: DC asks PM and TG about their winter break plans
- **Gangs2**: EC elicits a return to the gang discussion by PM
## Median PVI per topic

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Topic</th>
<th>n PVI</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PM</strong></td>
<td>Gangs1</td>
<td>397</td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td>Music</td>
<td>32</td>
<td>0.409</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>21</td>
<td>0.235</td>
</tr>
<tr>
<td></td>
<td>Movies</td>
<td>138</td>
<td>0.409</td>
</tr>
<tr>
<td></td>
<td>Break</td>
<td>63</td>
<td>0.250</td>
</tr>
<tr>
<td></td>
<td>Gangs2</td>
<td>9</td>
<td>0.347</td>
</tr>
</tbody>
</table>

| **TG**  | Gangs1     | 16    | 0.291  |
|         | Music      | 0     | 0.215  |
|         | Spanish    | 0     | 0.215  |
|         | Movies     | 155   | 0.360  |
|         | Break      | 149   | 0.270  |
|         | Gangs2     | 0     | 0.270  |

| **DC**  | Gangs1     | 312   | 0.445  |
|         | Music      | 47    | 0.255  |
|         | Spanish    | 0     | 0.255  |
|         | Movies     | 20    | 0.282  |
|         | Break      | 261   | 0.329  |
|         | Gangs2     | 0     | 0.329  |

| **EC**  | Gangs1     | 141   | 0.363  |
|         | Music      | 44    | 0.318  |
|         | Spanish    | 59    | 0.256  |
|         | Movies     | 66    | 0.317  |
|         | Break      | 0     | 0.317  |
|         | Gangs2     | 33    | 0.275  |

*Interviewees:*
- **Paco (PM)**
- **Tobi (TG)**

*Interviewers:*
- **Danica (DC)**
- **Erin (EC)**

- $p = .013$
- $p = .023$
Median PVI per topic

Interviewees: Paco (PM), Tobi (TG)
Interviewers: Danica (DC), Erin (EC)
Median PVI per topic

Interviewees: Paco (PM), Tobi (TG)
Interviewers: Danica (DC), Erin (EC)
Utterance length

- (In terms of number of PVI quotients, not temporal duration)

- Lab vs. naturalistic speech
  - Laboratory studies controlled for utterance length by having subjects read utterances of a set length
  - Utterances in naturalistic, spontaneous speech vary in length
Utterance PVI Values

- Higher PVI (Stress-timing)
  0.954 PVI, PM
  “And I was in the South Side”

- Lower PVI (Syllable-timing)
  0.120 PVI, TG
  “It’s about Hispanic people”
# Distribution of utterance lengths

<table>
<thead>
<tr>
<th>Speaker</th>
<th>n Utterances</th>
<th>Average Utterance Length (in PVI quotients)</th>
<th>% Ttl Utts with Utt Length 1 - 5 PVI quotients</th>
<th>% Ttl Utts with Utt Length ≥ 6 PVI quotients</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>685</td>
<td>4.23</td>
<td>75.3 %</td>
<td>24.7 %</td>
</tr>
<tr>
<td>TG</td>
<td>335</td>
<td>4.30</td>
<td>74.3 %</td>
<td>25.7 %</td>
</tr>
<tr>
<td>DC</td>
<td>684</td>
<td>6.77</td>
<td>58.0 %</td>
<td>42.0 %</td>
</tr>
<tr>
<td>EC</td>
<td>383</td>
<td>5.18</td>
<td>68.6 %</td>
<td>31.4 %</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>5.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Interviewees:**
- Paco (PM)
- Tobi (TG)

**Interviewers:**
- Danica (DC)
- Erin (EC)
**Utterance length**

**Means**

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Mean PVI Utt Length 1 - 5 PVI quotients</th>
<th>Mean PVI Utt Length ≥ 6 PVI quotients</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>0.416</td>
<td>0.443</td>
</tr>
<tr>
<td>TG</td>
<td>0.363</td>
<td>0.397</td>
</tr>
<tr>
<td>DC</td>
<td>0.418</td>
<td>0.470</td>
</tr>
<tr>
<td>EC</td>
<td>0.351</td>
<td>0.403</td>
</tr>
</tbody>
</table>

*Lower* means  
*Higher* means

(Mean of means per utterance-length)
A miniature corpus study

• Data from Pearsall, TX
  • Used in Thomas & Carter (2006), as well as by other research by P. Carter, T. Wolford, E. Callahan, etc.

• Both white and Hispanic interviewees
• Hispanic interviewees a mix of L1 and L2 English
Long vs. Short utterances
Results
(after observing variability)

• Speakers’ PVI will vary significantly depending on their language variety
  - Not in this data set

• A speaker’s PVI will not vary significantly depending on discursive factors
  - Actually, we do see variation within the discourse according to topic, segment, and utterance length
Overall Results

• Variation in PVI between speakers, but not according to language variety, as expected
• Variation in PVI within the discourse according to topic, segment, and utterance length

“Interesting though the resulting typological patterns [of languages rated with PVI scores] are, it is not at all clear what they are patterns of.” (Gibbon 2004)

• This metric is clearly measuring something, but what is it, exactly?
  – Variability in vowel duration
More than one type of “rhythm”? 

PVI has been said to be an index of speech rhythm. BUT, what if there are more than one independent dimensions we call “rhythm”?:

1. Related to syllable structure and segmental inventories
   - Basis for taxonomy
   - Variation within this domain is considered “noise”

2. Related to dynamic variations in speech fluency
   - Can vary within a single utterance
   - Harder to quantify
   - Subject to the “vagaries of expression and rhetorical force as much as to language-specific constraints”
   - The interaction between these two (hierarchically related) dimensions produces the perception of rhythm in speech

(Cummins 2004:2)
What do we mean by “rhythm”? 

- When looking within discourse, we might not mean the same thing we do when looking across languages
  - When a speaker’s PVI is higher during a particular topic section, for example, do we really want to say that they are speaking in a way that is more “English-like”?
  - Likewise, do we want to say that a short utterance is realized with more “Spanish-like” rhythm than is a long utterance?
What do we mean by “rhythm”? 

• May be premature to establish rhythmic typologies of languages until we can understand behavior of smaller units (Grabe 2002) 

“... the data would show whether the variable ‘language’ [TYPE 1] has an effect beyond that of the constituent ‘dialects’ [TYPE 1].” (Grabe 2002) 

AND 

“... the data would show whether the variable ‘dialect’ [TYPE 1] has an effect beyond that of the constituent ‘speakers’ [TYPE 2].” (Vaughn, just now)
Other approaches?

“...rhythm does not inhere in any language, but in the act of speaking.” (Cummins 2002)

- **Syllable/vowel durations** (could be epiphenomenal)
- What about?:
  - low-frequency Fourier analysis (Tilsen & Johnson 2007)
  - frequency/pitch
  - intensity/loudness
  - segmental quality (sonority scale)
  - syllable weight
  - foot/word/phrase-level stress patterns
    “Is rhythmicality an effect of accent, or is accent the effect of rhythmicality?” (Jassem, Hill & Witten 1984)
Thank you

Erik Thomas, Tyler Kendall,
Robin Dodsworth, Phillip Carter,
Erin Callahan, Danica Cullinan,
Walt Wolfram, Mary Kohn,
the North Carolina Language and Life Project,
NSF grants BCS-0213941 and BCS-0542139,

Janet Pierrehumbert & the Language Dynamics Lab
group, and Phonatics at Northwestern University
References


References, cont’d


